

0590  
0574

Page 1 of 5

AFH



OIPE

ENTERED

RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002

TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw

4 <110> APPLICANT: Dunn-Coleman, Nigel  
5 Goedegebuur, Frits  
6 Ward, Michael  
7 Yao, Jian  
9 <120> TITLE OF INVENTION: EGVIII Endoglucanase and Nucleic Acids  
10 Encoding the Same  
12 <130> FILE REFERENCE: GC700  
14 <140> CURRENT APPLICATION NUMBER: US 10/028,245  
15 <141> CURRENT FILING DATE: 2001-12-18  
17 <160> NUMBER OF SEQ ID NOS: 4  
19 <170> SOFTWARE: FastSEQ for Windows Version 4.0  
21 <210> SEQ ID NO: 1  
22 <211> LENGTH: 1826  
23 <212> TYPE: DNA  
24 <213> ORGANISM: Trichoderma reesei  
26 <400> SEQUENCE: 1  
27 gtcgaccac gctccgttc attcttcctc ccctcctcct ctcctccctc ctccccttct 60  
28 cccccatcac cgtcaccact ctccctcattt cgcctctctc tgcgagccat gacgcagcat 120  
29 taacacacac tcgttctgt tactctcgct gtcgtcgct ctgctcggt gcattctgct 180  
30 tagcactttt tttcgctctt cgttctcttt taatccgtca tcttctgcaa tctgctgcca 240  
31 tttgttcgac taggttagtgg taatatacgg acagctttt ttcctcgct caacacgtcg 300  
32 acgtacaatt aatacaccat ctcgttaatc ggatatatcc ctcggccctct tcctgggtct 360  
33 tgtgcgacgc tcctcggttc tccctctcat tatgcgcgca acctcccttc tggccgccc 420  
34 cttggccgtg gctggcgatg ccctcgccgg caagatcaaa tatctggcg tcgccattcc 480  
35 cggaatcgac tttggctgacg acatcgacgg cagctgtccg actgacacgt cgtctgtgcc 540  
36 cctgctgagc tacaaaggag gagatggcgc cggccagatg aagcatttcg cggaaagacga 600  
37 cggcctcaac gtcttcgca tatccgctac atggcagttt gtcctcaaca acacgggtgga 660  
38 cggcaagctg gacgagctca actggggctc ctacaacaag gtcgtcaacg cctgtctcga 720  
39 gacgggcgcc tactgcatga ttgacatgca caacttgcg cgctacaacg gcccgtatcat 780  
40 cggccaggaa ggcgtgtcgg acgacatctt tgtcgaccc tgggtccaga tcgcaaagta 840  
41 ctacaggagc aacgacaaga tcatctttgg cctgatgaac gagccgcacg acctcgacat 900  
42 ttagatctgg ggcagacgt gccaaaagggt cgtcaactgacg atccgaaagg ccggccac 960  
43 ctcgcagatg atcctcctgc ccggaaaccaa ctttgcacg gtcgagacgt atgtgtccac 1020  
44 tggcagcgcg gaagccctcg gcaagattac gaaccggat ggaaggcaccg atttgtgtta 1080  
45 ctttgcgttc cacaagtatac tgcacatcaa caactccggg tcgcacgcgg agtgcaccac 1140  
46 agacaacgtc gacgccttca acgacttcgc ggactggctg aggcagaaca agcggcaggc 1200  
47 catcatctcc gaaacggcg cgtccatgga accttcgtgc atgactgcct tctgcgccc 1260  
48 gaacaaggcc attagcgaaa acagcgacgt ctacattggc tttgtggct ggggtccgg 1320  
49 cagcttgac acgtcgatca tcttgactct gactccctc ggcagcccg gcaactacac 1380  
50 cgacacaacaag ctcatgaacg agtgcattct ggaccagttt accctcgacg aaaagtaccg 1440  
51 tccaaacaccc acctcaattt ccacagccgc ggaagagacg gccacggcga cagcaacctc 1500  
52 tgacggcgac ggcgcattca ctacgaagcc catctttagg gaagaaaccg cctctccac 1560  
53 tcccaatgct gttaccaagc cctcgcccgac cacgagcgac tcttccgacg acgacaagga 1620

RAW SEQUENCE LISTING  
PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002  
TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt  
Output Set: N:\CRF3\05102002\J028245.raw

54 ctcggcagca tctatgagtg cccaggcctt gacaggcacg gtgctgttta ctgttgctgc 1680  
 55 ccttggctac atgctggtag cgttttgatg tttttttttt aatgagtttg tatacctaatt 1740  
 56 gagcatgatt gagatgctac gtatgtatata tgtctttacg ggtacataaag actagagccaa 1800  
 57 tggtaatc aaaaaaaaaaaaaaa 1826  
 59 <210> SEQ ID NO: 2  
 60 <211> LENGTH: 419  
 61 <212> TYPE: PRT  
 62 <213> ORGANISM: Trichoderma reesei  
 64 <400> SEQUENCE: 2  
 65 Gly Lys Ile Lys Tyr Leu Gly Val Ala Ile Pro Gly Ile Asp Phe Gly  
 66 1 5 10 15  
 67 Cys Asp Ile Asp Gly Ser Cys Pro Thr Asp Thr Ser Ser Val Pro Leu  
 68 20 25 30  
 69 Leu Ser Tyr Lys Gly Gly Asp Gly Ala Gly Gln Met Lys His Phe Ala  
 70 35 40 45  
 71 Glu Asp Asp Gly Leu Asn Val Phe Arg Ile Ser Ala Thr Trp Gln Phe  
 72 50 55 60  
 73 Val Leu Asn Asn Thr Val Asp Gly Lys Leu Asp Glu Leu Asn Trp Gly  
 74 65 70 75 80  
 75 Ser Tyr Asn Lys Val Val Asn Ala Cys Leu Glu Thr Gly Ala Tyr Cys  
 76 85 90 95  
 77 Met Ile Asp Met His Asn Phe Ala Arg Tyr Asn Gly Gly Ile Ile Gly  
 78 100 105 110  
 79 Gln Gly Gly Val Ser Asp Asp Ile Phe Val Asp Leu Trp Val Gln Ile  
 80 115 120 125  
 81 Ala Lys Tyr Tyr Glu Asp Asn Asp Lys Ile Ile Phe Gly Leu Met Asn  
 82 130 135 140  
 83 Glu Pro His Asp Leu Asp Ile Glu Ile Trp Ala Gln Thr Cys Gln Lys  
 84 145 150 155 160  
 85 Val Val Thr Ala Ile Arg Lys Ala Gly Ala Thr Ser Gln Met Ile Leu  
 86 165 170 175  
 87 Leu Pro Gly Thr Asn Phe Ala Ser Val Glu Thr Tyr Val Ser Thr Gly  
 88 180 185 190  
 89 Ser Ala Glu Ala Leu Gly Lys Ile Thr Asn Pro Asp Gly Ser Thr Asp  
 90 195 200 205  
 91 Leu Leu Tyr Phe Asp Val His Lys Tyr Leu Asp Ile Asn Asn Ser Gly  
 92 210 215 220  
 93 Ser His Ala Glu Cys Thr Thr Asp Asn Val Asp Ala Phe Asn Asp Phe  
 94 225 230 235 240  
 95 Ala Asp Trp Leu Arg Gln Asn Lys Arg Gln Ala Ile Ile Ser Glu Thr  
 96 245 250 255  
 97 Gly Ala Ser Met Glu Pro Ser Cys Met Thr Ala Phe Cys Ala Gln Asn  
 98 260 265 270  
 99 Lys Ala Ile Ser Glu Asn Ser Asp Val Tyr Ile Gly Phe Val Gly Trp  
 100 275 280 285  
 101 Gly Ala Gly Ser Phe Asp Thr Ser Tyr Ile Leu Thr Leu Thr Pro Leu  
 102 290 295 300  
 103 Gly Lys Pro Gly Asn Tyr Thr Asp Asn Lys Leu Met Asn Glu Cys Ile  
 104 305 310 315 320

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002

TIME: 09:51:18

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw

105 Leu Asp Gln Phe Thr Leu Asp Glu Lys Tyr Arg Pro Thr Pro Thr Ser  
 106 325 330 335  
 107 Ile Ser Thr Ala Ala Glu Glu Thr Ala Thr Ala Thr Ser Asp  
 108 340 345 350  
 109 Gly Asp Ala Pro Ser Thr Thr Lys Pro Ile Phe Arg Glu Glu Thr Ala  
 110 355 360 365  
 111 Ser Pro Thr Pro Asn Ala Val Thr Lys Pro Ser Pro Asp Thr Ser Asp  
 112 370 375 380  
 113 Ser Ser Asp Asp Asp Lys Asp Ser Ala Ala Ser Met Ser Ala Gln Gly  
 114 385 390 395 400  
 115 Leu Thr Gly Thr Val Leu Phe Thr Val Ala Ala Leu Gly Tyr Met Leu  
 116 405 410 415  
 117 Val Ala Phe  
 120 <210> SEQ ID NO: 3  
 121 <211> LENGTH: 19  
 122 <212> TYPE: PRT  
 123 <213> ORGANISM: Trichoderma reesei  
 125 <400> SEQUENCE: 3  
 126 Met Arg Ala Thr Ser Leu Leu Ala Ala Ala Leu Ala Val Ala Gly Asp  
 127 1 5 10 15  
 128 Ala Leu Ala  
 131 <210> SEQ ID NO: 4  
 132 <211> LENGTH: 1317  
 133 <212> TYPE: DNA  
 134 <213> ORGANISM: Trichoderma reesei  
 136 <400> SEQUENCE: 4  
 137 atgcgcgcaa ccccttctt ggcggccggc ttggccgtgg ctggcgatgc cctcgccggc 60  
 138 aagatcaaat atctggcggt cgccattccc ggaatcgact ttggctgcga catcgacggc 120  
 139 agctgtccga ctgacacgtc gtctgtgcgg ctgtctgagct acaaaggagg agatggcgcc 180  
 140 ggccagatga agcatttcgc cgaagacgac ggccctcaacg tctttcgcat atccgctaca 240  
 141 tggcagtttgc tcctcaacaa cacgggtggac ggcaagctgg acgagctcaa ctggggctcc 300  
 142 tacaacaagg tcgtcaacgc ctgtctcgag acggcgccct actgcattat tgacatgcac 360  
 143 aactttggcc gctacaacgg cggcatcatc ggccaggagg gctgttcgga cgacatctt 420  
 144 gtcgacctctt gggccatcgat cgcacaaatgac tacgaggaca acgacaagat catctttggc 480  
 145 ctgtatgaacg agccgcacga cctcgacatt gagatctggg cgcagacgtg ccaaaagggtc 540  
 146 gtcactgcga tccgaaaggc cggcgccacc tcgcagatga tcctctgtcc cgaaaccaac 600  
 147 ttggccagcg tcgagacgta tttgtccact ggccggccgg aagccctcgaa caagattacg 660  
 148 aacccggatg gaagcaccga tttgtctgtac tttgtatgtcc acaagttatct cgacatcaac 720  
 149 aactccgggt cgcaccccgaa gtgcaccacca gacaacgtcg acgccttcaa cgacttcgcg 780  
 150 gactggctga ggcagaacaa ggcggccacc atcatctccg aaacggggcgc gtccatggaa 840  
 151 ctttcgtgca tgactgcctt ctgcggcccg aacaaggcca ttgcgaaaa cagcgacgtc 900  
 152 tacattggct ttgtggctg gggccggc agcttgcata cgctgtacat cttgactctg 960  
 153 actccccctcg gcaaggcccgaa caactacacc gacaacaaggc tcatgaacga gtgcattctg 1020  
 154 gaccagtttccctcgacgaa aagttacccgtt ccaacacccaa cctcaatttc cacagcggcg 1080  
 155 gaagagacgg ccacggcgac agcaacccctt gacggcgacg cggccatccac tacgaagccc 1140  
 156 atcttttaggg aagaaaccgc ctctcccact cccaaatgtt ttaccaagcc ctcggccgac 1200  
 157 acgagcgact cttccgacgaa cgacaaggac tcggcagcat ctatgagtgc ccagggttt 1260  
 158 acaggcacgg tgctgtttac ttggctgtcc cttggctaca tgctggtagc gttttga 1317

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/10/028,245

DATE: 05/13/2002

TIME: 09:51:19

Input Set : A:\GC700-SEQLIST.txt

Output Set: N:\CRF3\05102002\J028245.raw